

Explicit Constructions of Generalized MSTD Sets

Steven J. Miller, Sean Pegado and Luc Robinson
Williams College

http://www.williams.edu/Mathematics/sjmiller/public_html

CANT 2011

Summary

- History of the problem.
- Examples.
- Main results and proofs.
- Open problems and conjectures.

Introduction

Statement

Let A be a finite set of integers, $|A|$ its size. Form

- Sumset: $A + A = \{a_i + a_j : a_i, a_j \in A\}$.
- Difference set: $A - A = \{a_i - a_j : a_i, a_j \in A\}$.

Statement

Let A be a finite set of integers, $|A|$ its size. Form

- Sumset: $A + A = \{a_i + a_j : a_i, a_j \in A\}$.
- Difference set: $A - A = \{a_i - a_j : a_i, a_j \in A\}$.

Definition

We say A is **difference dominated** if $|A - A| > |A + A|$, **balanced** if $|A - A| = |A + A|$ and **sum dominated (or an MSTD set)** if $|A + A| > |A - A|$.

Questions

We expect **generic** set to be difference dominated:

- Addition is commutative, subtraction isn't.
- Generic pair (x, y) gives 1 sum, 2 differences.

Questions

We expect **generic** set to be difference dominated:

- Addition is commutative, subtraction isn't.
- Generic pair (x, y) gives 1 sum, 2 differences.

Questions

- Do there exist sum-dominated sets?
- If yes, how many?

Examples

Examples

- Conway: $\{0, 2, 3, 4, 7, 11, 12, 14\}$.
- Marica (1969): $\{0, 1, 2, 4, 7, 8, 12, 14, 15\}$.
- Freiman and Pigarev (1973): $\{0, 1, 2, 4, 5, 9, 12, 13, 14, 16, 17, 21, 24, 25, 26, 28, 29\}$.
- Computer search of random subsets of $\{1, \dots, 100\}$:
 $\{2, 6, 7, 9, 13, 14, 16, 18, 19, 22, 23, 25, 30, 31, 33, 37, 39, 41, 42, 45, 46, 47, 48, 49, 51, 52, 54, 57, 58, 59, 61, 64, 65, 66, 67, 68, 72, 73, 74, 75, 81, 83, 84, 87, 88, 91, 93, 94, 95, 98, 100\}$.
- Recently infinite families (Hegarty, Nathanson).

Infinite Families

Key observation

If A is an arithmetic progression, $|A + A| = |A - A|$.

Proof:

- WLOG, $A = \{0, 1, \dots, n\}$ as $A \rightarrow \alpha A + \beta$ doesn't change $|A + A|$, $|A - A|$.
- $A + A = \{0, \dots, 2n\}$, $A - A = \{-n, \dots, n\}$, both of size $2n + 1$. □

Previous Constructions

Many constructions perturb an arithmetic progression.

Example:

- MSTD set $A = \{0, 2, 3, 4, 7, 11, 12, 14\}$.
- $A = \{0, 2\} \cup \{3, 7, 11\} \cup (14 - \{0, 2\}) \cup \{4\}$.

Notation

- Define $[a, b] = \{k \in \mathbb{Z} : a \leq k \leq b\}$.
- A is a P_n^2 -set if its sumset $A + A$ and difference set $A - A$ contain all but the first and last n possible elements (may or may not contain some of fringe elements).
- A P_n^4 -set is a set where $A + A + A + A$ and $A + A - A - A$ each contain all but the first and last n elements.

Large Explicit Family

Theorem (Miller-Orosz-Scheinerman)

$A = L \cup R$ an MSTD, P_n^2 -set, O_k is k consecutive elements, M no runs of k missing elements, then $A' = L \cup O_k \cup M \cup O'_k \cup R'$ an MSTD set and at least $C2^r/r^4$ of subsets of $\{0, \dots, r-1\}$ are MSTD.

Large Explicit Family

Theorem (Miller-Orosz-Scheinerman)

$A = L \cup R$ an MSTD, P_n^2 -set, O_k is k consecutive elements, M no runs of k missing elements, then $A' = L \cup O_k \cup M \cup O'_k \cup R'$ an MSTD set and at least $C2^r/r^4$ of subsets of $\{0, \dots, r-1\}$ are MSTD.

Proof: If M never misses k consecutive elements, have all possible sums/differences.

Large Explicit Family

Theorem (Miller-Orosz-Scheinerman)

$A = L \cup R$ an MSTD, P_n^2 -set, O_k is k consecutive elements, M no runs of k missing elements, then $A' = L \cup O_k \cup M \cup O'_k \cup R'$ an MSTD set and at least $C2^r/r^4$ of subsets of $\{0, \dots, r-1\}$ are MSTD.

Proof: If M never misses k consecutive elements, have all possible sums/differences.

OK if have at least one in blocks of $k/2$ elements, yields percentage in $\{0, \dots, r-1\}$ is

$$\gg \sum_{k=n}^{r/4} \frac{1}{2^{2k}} \left(1 - \frac{1}{2^{k/2}}\right)^{\frac{r}{k/2}}.$$

Large Explicit Family

Theorem (Miller-Orosz-Scheinerman)

$A = L \cup R$ an MSTD, P_n^2 -set, O_k is k consecutive elements, M no runs of k missing elements, then $A' = L \cup O_k \cup M \cup O'_k \cup R'$ an MSTD set and at least $C2^r/r^4$ of subsets of $\{0, \dots, r-1\}$ are MSTD.

Proof: If M never misses k consecutive elements, have all possible sums/differences.

OK if have at least one in blocks of $k/2$ elements, yields percentage in $\{0, \dots, r-1\}$ is

$$\gg \sum_{k=n}^{r/4} \frac{1}{2^{2k}} \left(1 - \frac{1}{2^{k/2}}\right)^{\frac{r}{k/2}}.$$

O_k and O'_k completely determined. If k grows with n , cannot have a positive percentage.

Results

Constructing Sets with $|A + A + A + A| > |A + A - A - A|$

- Begin by searching for a single set with $|A + A + A + A| > |A + A - A - A|$.
- Generate random subsets of $[1, 40]$, include each number with probability $1/4$, and check if the generated set has the desired property.
- Two million trials yield

$$\{6, 7, 9, 10, 13, 32, 35, 36, 38, 39, 40\},$$

which has $|A + A + A + A| = 136$ and $|A + A - A - A| = 135$.

Constructing Infinite Families $|A + A + A + A| > |A + A - A - A|$

To use one set to create a large infinite family of sets we modified Miller, Orosz, and Scheinerman's construction. The set must satisfy two properties:

- The set is a subset of $[1, 2n]$ containing 1 and $2n$.
- The set must be a P_n^4 set.

Some work yields

$$A = \{1, 2, 4, 5, 8, 27, 30, 31, 33, 34, 35, 50, \\ 51, 53, 54, 57, 76, 79, 80, 82, 83, 84\} \subset [1, 2n]$$

with these desired properties.

Constructing Infinite Families, Continued

Our new set allows us to prove the following result:

Theorem (MPR)

There is a constant $C > 0$ such that as n goes to infinity, the percentage of subsets A of $[1, n]$ with $|2A + 2A| > |2A - 2A|$ is at least $C/n^{4/3}$.

Proof: Modify method of Miller-Orosz-Scheinerman.

Better Proof: Use results on length of consecutive heads in coin tosses and replace $4/3$ with $2/3$.

Constructing Infinite Families, Improved

Theorem (MPR)

There is a constant $C > 0$ such that as n goes to infinity, the percentage of subsets A of $[1, n]$ with $|2A + 2A| > |2A - 2A|$ is at least C/n^r , where $r = \frac{1}{6} \log_2(256/255) < .001$.

Constructing Infinite Families, Improved

Theorem (MPR)

There is a constant $C > 0$ such that as n goes to infinity, the percentage of subsets A of $[1, n]$ with $|2A + 2A| > |2A - 2A|$ is at least C/n^r , where $r = \frac{1}{6} \log_2(256/255) < .001$.

Proof: $A = L \cup R$, consider $A' = L \cup O_1 \cup M \cup O_2 \cup R'$.

- O 's show up in sums/differences at least in pairs, unless with $L + L + L$, $R' + R' + R'$ or $L + L - R'$.

Constructing Infinite Families, Improved

Theorem (MPR)

There is a constant $C > 0$ such that as n goes to infinity, the percentage of subsets A of $[1, n]$ with $|2A + 2A| > |2A - 2A|$ is at least C/n^r , where $r = \frac{1}{6} \log_2(256/255) < .001$.

Proof: $A = L \cup R$, consider $A' = L \cup O_1 \cup M \cup O_2 \cup R'$.

- O 's show up in sums/differences at least in pairs, unless with $L + L + L$, $R' + R' + R'$ or $L + L - R'$.
- Each of $L + L + L$, $R' + R' + R'$ and $L + L - R'$ contain a run of 16 elements in a row.

Constructing Infinite Families, Improved

Theorem (MPR)

There is a constant $C > 0$ such that as n goes to infinity, the percentage of subsets A of $[1, n]$ with $|2A + 2A| > |2A - 2A|$ is at least C/n^r , where $r = \frac{1}{6} \log_2(256/255) < .001$.

Proof: $A = L \cup R$, consider $A' = L \cup O_1 \cup M \cup O_2 \cup R'$.

- O 's show up in sums/differences at least in pairs, unless with $L + L + L$, $R' + R' + R'$ or $L + L - R'$.
- Each of $L + L + L$, $R' + R' + R'$ and $L + L - R'$ contain a run of 16 elements in a row.
- Can relax the MOS structure (each O was k consecutive elements); if each O has no run of 16 missing elements and $2O$ full for both O 's, get all sums/differences.

Constructing Infinite Families, Improved

Theorem (MPR)

There is a constant $C > 0$ such that as n goes to infinity, the percentage of subsets A of $[1, n]$ with $|2A + 2A| > |2A - 2A|$ is at least C/n^r , where $r = \frac{1}{6} \log_2(256/255) < .001$.

Proof: $A = L \cup R$, consider $A' = L \cup O_1 \cup M \cup O_2 \cup R'$.

- O 's show up in sums/differences at least in pairs, unless with $L + L + L$, $R' + R' + R'$ or $L + L - R'$.
- Each of $L + L + L$, $R' + R' + R'$ and $L + L - R'$ contain a run of 16 elements in a row.
- Can relax the MOS structure (each O was k consecutive elements); if each O has no run of 16 missing elements and $2O$ full for both O 's, get all sums/differences.
- Replace the $1/2^{2k}$ from the O s in MOS with a better exponent.

Selecting the Difference Between Sumset and Difference Set

We can construct a set S_x such that the difference between the sumset and the difference set of S_x is x for any integer x .

Theorem (MPR)

Given $x \in \mathbb{Z}$ there exists $S_x \subset [0, 157|x|]$ with $|2S_x + 2S_x| - |2S_x - 2S_x| = x$. For large x , there exists $S_x \subset [0, 35|x|]$.

Proof: Start with S_1 with $x = 1$. Shift and amalgamate gives S_5 , and repeat with shifts of S_1 and get S_x for $x \equiv 1 \pmod{4}$.

Cull the S_x 's to move backwards to get missing $x \pmod{4}$.

Open Problems

When is a Set Sum-Dominated?

Initial observations: sets including 0 and primes up to p for small primes p are difference-dominated.

- But also, typically $|A + A| - |A - A| = |2A + 2A| - |2A - 2A|$
 $= |4A + 4A| - |4A - 4A| = |8A + 8A| - |8A - 8A| =$
 $|16A + 16A| - |16A - 16A|$, and continued to hold in higher
sums including $|1024A + 1024A|$.

When is a Set Sum-Dominated?

Initial observations: sets including 0 and primes up to p for small primes p are difference-dominated.

- But also, typically $|A + A| - |A - A| = |2A + 2A| - |2A - 2A| = |4A + 4A| - |4A - 4A| = |8A + 8A| - |8A - 8A| = |16A + 16A| - |16A - 16A|$, and continued to hold in higher sums including $|1024A + 1024A|$.
- True for all checked sets containing 0 and $\{\text{primes} + 1\}$ and even random (normalized sets) tested.

When is a Set Sum-Dominated?

Initial observations: sets including 0 and primes up to p for small primes p are difference-dominated.

- But also, typically $|A + A| - |A - A| = |2A + 2A| - |2A - 2A| = |4A + 4A| - |4A - 4A| = |8A + 8A| - |8A - 8A| = |16A + 16A| - |16A - 16A|$, and continued to hold in higher sums including $|1024A + 1024A|$.
- True for all checked sets containing 0 and $\{\text{primes} + 1\}$ and even random (normalized sets) tested.
- Only occasional exception: $|A + A| - |A - A|$ sometimes different. All other differences are equal.

When is a Set Sum-Dominated?

A	A+A	A-A	A+A+A+A	A+A-A-A	4A+4A	4A-4A	8A+8A	8A-8A	16A+16A	16A-16A
primes to 7	12	13	26	27	54	55	110	111	222	223
to 11	17	21	39	43	83	87	171	175	347	351
to 13	21	25	47	51	99	103	203	207	411	415
to 17	26	33	60	67	128	135	264	271	536	543
to 19	30	37	66	75	144	151	296	303	600	607
to 23	35	45	81	91	173	183	357	367	725	735
to 29	40	55	98	115	214	231	446	463	910	927
to 31	46	59	110	123	234	247	482	495	978	991
to 37	52	69	127	147	275	295	571	591	1163	1183
Looking at primes off by 1										
{0,3,4,6}	9	11	21	23	45	47	93	95	189	191
{0,3,4,6,8}	12	15	28	31	60	63	124	127	252	255
{0,3,4,6,8,12}	16	19	40	43	88	91	184	187	376	378
{0,3,4,6,8,12,14}	20	25	49	55	105	111	217	223	441	447
{0,3,4,6,8,12,14,16}	24	29	61	67	133	139	277	283	565	571
{0,3,4,6,8,12,14,16,20}	28	35	70	79	150	159	310	319	630	639
{0,3,4,6,8,12,14,16,20,24}	32	39	82	91	178	187	370	379	754	763
{0,3,4,6,8,12,14,16,20,24,28}	36	47	97	115	217	235	457	475	937	955
{0,3,4,6,8,12,14,16,20,24,28,32}	42	53	112	127	240	255	496	511	1008	1023
{0,3,4,6,8,12,14,16,20,24,28,32,36}	47	59	127	147	279	299	583	603	1191	1211
(Normalized) Random										
{0,4,5,6}	9	11	21	25	45	49	93	97	189	193
{0,2,3,5,7,6}	16	17	32	33	64	65	128	129	256	257
{0,2,3,5,6,9,10}	20	21	40	41	80	81	160	161	320	321
{0,1,2,3,6,13,16}	22	29	58	65	122	129	250	257	506	513
{0,3,5,9,12,13,15,1}	30	33	74	77	158	161	326	329	662	665
{0,3,5,7,11,12,17,1}	41	49	99	103	207	211	423	427	855	859
{0,3,4,5,6,7,8,12,1}	30	35	66	71	138	143	282	287	570	575

Questions Raised

- Does this property hold for all sets with probability 1?
- Originally, this looked to be something unique to the sets of primes, but that was not the case. Are sets including primes interesting or unique in any other way?